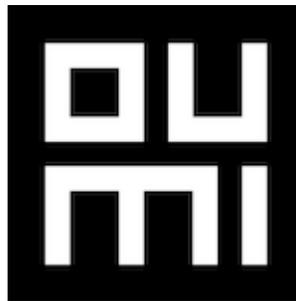




Getting Started

Contents

- Getting Started
- Why use Oumi?
- Where to go next?
- Join the Community!
- Need Help?



[Github oumi](#)
[Blog oumi](#)
[Discord 107 online](#)
[pypi package 0.1.3](#)
[License Apache 2.0](#)
[Stars 6.9k](#)

[code style black](#)
[pre-commit enabled](#)
[About oumi](#)

Everything you need to build state-of-the-art foundation models, end-to-end.



Oumi is a fully open-source platform that streamlines the entire lifecycle of foundation models - from data preparation and training to evaluation and deployment. Whether you're developing on a laptop, launching large scale experiments on a cluster, or deploying models in production, Oumi provides the tools and workflows you need.

With Oumi, you can:

-  Train and fine-tune models from 10M to 405B parameters using state-of-the-art techniques (SFT, LoRA, QLoRA, DPO, and more)
-  Work with both text and multimodal models (Llama, DeepSeek, Qwen, Phi, and others)
-  Synthesize and curate training data with LLM judges
-  Deploy models efficiently with popular inference engines (vLLM, SGLang)
-  Evaluate models comprehensively across standard benchmarks
-  Run anywhere - from laptops to clusters to clouds (AWS, Azure, GCP, Lambda, and more)
-  Integrate with both open models and commercial APIs (OpenAI, Anthropic, Vertex AI, Parasail, ...)

All with one consistent API, production-grade reliability, and all the flexibility you need for research. Oumi is currently in [beta](#) and under active development.

Notebook	Try in Colab	Goal
 Getting Started: A Tour	 Open in Colab	Quick tour of core features: training, evaluation, inference, and job management
 Model Finetuning Guide	 Open in Colab	End-to-end guide to LoRA tuning with data prep, training, and evaluation
 Model Distillation	 Open in Colab	Guide to distilling large models into smaller, efficient ones
 Model Evaluation	 Open in Colab	Comprehensive model evaluation using Oumi's evaluation framework
 Remote Training	 Open in Colab	Launch and monitor training jobs on cloud (AWS, Azure, GCP, Lambda, etc.) platforms
 LLM-as-a-Judge	 Open in Colab	Filter and curate training data with built-in judges
 vLLM Inference Engine	 Open in Colab	Fast inference at scale with the vLLM engine

If you need a comprehensive platform for training, evaluating, or deploying models, Oumi is a great choice.